




EXHIBIT B

REDACTED

Search Terms for Email and Workplace Chat

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
[For Aurelien Rodriguez and Hugo Touvron] (pre-train* OR pretrain* OR dataset* OR "data set*" OR "training data") AND (llama OR genesis OR "large language model" OR LLM)	51,000	57,279
((Llama OR Genesis) w/15 "open source") OR ((Llama OR Genesis) w/5 (open AND (release OR license)))	7,985	12,050
(Copyright* w/15 compensat*) OR (Copyright* w/15 consent*) OR (Copyright* w/15 licens*) OR (copyright* w/15 credit*) OR (copyright* w/15 attribut*) OR (author* w/15 compensat*) OR (author* w/15 consent*) OR (author* w/15 licens*) OR (author* w/15 credit*) OR (author* w/15 attribut*)	4,490	16,667
((Copyright* w/15 compensat*) OR (Copyright* w/15 consent*) OR (Copyright* w/15 licens*) OR (copyright* w/15 credit*) OR (copyright* w/15 attribut*) OR (author* w/15 compensat*) OR (author* w/15 consent*) OR (author* w/15 licens*) OR (author* w/15 credit*) OR (author* w/15 attribut*)) AND (LLM OR Llama OR GenAI OR train* OR "large language model")	4,138	16,269
(market* w/25 Llama) OR (market* w/25 Genesis) OR (promo* w/25 Llama) OR (promo* w/25 Genesis)	3,390	7,413
████████████████████ ████████████████████ ████████████████████ ████████████████████ ████████████████████ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	2,937	9,915
Books3 OR "books 3" OR b3g OR Gutenberg OR Libgen OR "lib gen" OR "Library genesis"	2,785	9,307
████████████████████ ████████████████████ ████████████████████ ████████████████████ ████████████████████ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	1,773	5,735
████████████████████ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	1,673	6,065
libgen OR "library genesis" OR "anna's archive" OR "annas archive" OR bibliotik OR zlibrary OR z-library	1,494	2,348
Books3 OR "books 3" OR b3g OR Gutenberg	1,448	7,381
(go-to-market OR "go to market" OR "executive summary" OR roadmap OR "road map" OR SWOT) w/20 (Llama OR Genesis)	1,129	2,192

(Schumer w/10 AI) OR (Schumer w/10 Artificial) OR (senat* w/10 AI) OR (senat* w/10 Artificial)	834	1,118
(mitigat* w/20 memoriz*) OR (mitigat* w/20 copyright) OR (mitigat* w/20 IP) OR (mitigat* w/20 "intellectual property")	788	1,255
[For all ESI custodians other than Aurelien Rodriguez and Hugo Touvron] (pre-train* OR pretrain* OR dataset* OR "data set*" OR "training data") AND [Header includes:] ("Library Genesis" OR LibGen OR "lib gen" OR libgen.li OR Z-Library OR "z library" OR zlibrary OR B-ok OR b-ok.cc OR zlib OR zlib.io OR Sci-Hub OR scihub OR sci-hub.pub OR Bibliotik OR bibliotik.me OR "the eye" OR the-eye.eu OR "Anna's Archive" OR annas-archive.org OR shawn@keenagi.com OR shawnpresser@gmail.com OR Stella OR Biderman OR eleuthero.ai OR Eleuther OR (leo w/3 gao) OR leogao* OR microsoft OR microsoft.com OR OpenAI OR openai.com OR anthropic OR anthropic.com OR Dell OR dell.com OR Cloudflare OR cloudflare.com OR Qualcomm OR qualcomm.com OR timdettmers OR Dettmers"	567	908
   w/10 (licens* OR contract* OR agreement* OR negotiat* OR partner*)) AND (LLM OR Llama OR GenAI OR train* OR "large language model")	456	3,194
((("meta AI" w/5 Facebook) OR ("meta AI" w/5 Instagram) OR ("meta AI" w/5 Reels) OR ("meta AI" w/5 whatsapp) OR ("meta AI" w/5 Threads) OR ("meta AI" w/5 oculus) OR ("meta AI" w/5 metaverse) OR ("meta AI" w/5 orion) OR ("meta AI" w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization)	410	817
(licens* w/3 book*) OR (licens* w/3 textbook*) OR (licens* w/3 novel*) OR (licens* w/3 literature)	403	2,672
(Software OR tool) w/20 ((Copyright* w/15 compensat*) OR (Copyright* w/15 consent*) OR (Copyright* w/15 licens*) OR (copyright* w/15 credit*) OR (copyright* w/15 attribut*) OR (author* w/15 compensat*) OR (author* w/15 consent*) OR (author* w/15 licens*) OR (author* w/15 credit*) OR (author* w/15 attribut*))	346	1,120
(litigation OR lawsuit OR "fair use") w/20 (Kadrey OR Silverman OR Golden OR Woodson OR Greer OR Snyder OR Hwang OR Coates OR Lippman OR Klam OR Diaz OR Díaz OR Terkeurst OR Farnsworth)	318	495

(licens* w/3 books) OR (licens* w/3 textbooks) OR (licens* w/3 novels) OR (licens* w/3 literature)	300	2,460
((Compet* w/5 analys*) OR (Compet* w/5 assess*)) AND (Llama OR Genesis)	255	729
"publicly available datasets" AND llama	229	1,061
(Llama w/10 budget) OR (Llama w/10 "operational expenses") OR (Llama w/10 "gross margin") OR (Llama w/10 "operating margin") OR (Llama w/10 profit) OR (Llama w/10 EPS) OR (Llama w/10 "earnings per share")	222	486
((Llama w/5 Facebook) OR (Llama w/5 Instagram) OR (Llama w/5 Reels) OR (Llama w/5 whatsapp) OR (Llama w/5 Threads) OR (Llama w/5 oculus) OR (Llama w/5 metaverse) OR (Llama w/5 orion) OR (Llama w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization)	139	464
((book* w/20 storytelling) OR (book* w/20 narrative) OR (book* w/20 value) OR (book* w/20 benefit)) AND (dataset OR "training data")	116	907
[REDACTED] w/10 (licens* OR contract* OR agreement* OR negotiat* OR partner*)) AND (LLM OR Llama OR GenAI OR train* OR "large language model")	108	171
"AI Insight Forum"	101	126
"theorem proving" AND (dataset OR "training data")	100	522
torrent AND (dataset OR "training data")	97	434
("business plan" OR "marketing plan" OR "growth plan" OR (business* w/3 summar*) OR (market* w/3 summar*) OR (grow* w/3 summar*) OR (strateg* w/3 summar*)) w/50 Llama	87	166
((Shawn w/3 Presser) OR shawnpresser) w/50 (pre-train* OR pretrain* OR dataset* OR "data set*" OR "training data" OR "the pile" OR Llama OR Genesis)	83	616
[REDACTED] w/10 (licens* OR contract* OR agreement* OR negotiat* OR partner*)) AND (LLM OR Llama OR GenAI OR train* OR "large language model")	73	176

((Richard w/3 Kadrey) OR (Sarah w/3 Silverman) OR (Christopher w/3 Golden) OR (Jacqueline w/3 Woodson) OR (Andrew w/3 Greer) OR (Andrew w/3 Sean)) w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	70	108
(Contract* w/30 RLHF) OR (Contract* w/30 annotation*) OR (Agreement* w/30 RLHF) OR (Agreement* w/30 annotation*) OR (statement w/30 RLHF) OR (SOW w/30 RLHF) OR (statement w/30 annotation*) OR (SOW w/30 annotation*) OR (SOW w/30 RLHF)	53	69
((Sean w/3 Greer) OR (Rachel w/3 Snyder) OR (David w/3 Hwang) OR (Ta-Nehisi w/3 Coates) OR (Laura w/3 Lippman)) w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	30	74
(christopher w/3 farnsworth) OR "blood oath" OR "president's vampire" OR "red, white, and blood" OR "red, white, & blood" OR "a nathaniel cade story" OR "the burning men" OR "eternal world" OR killfile OR "Deep state: a nathaniel cade story" OR ("deep state" AND farnsworth) OR (flashmob AND farnsworth)	22	2,341
("Meta AI" w/10 budget) OR ("Meta AI" w/10 "operational expenses") OR ("Meta AI" w/10 "gross margin") OR ("Meta AI" w/10 "operating margin") OR ("Meta AI" w/10 profit) OR ("Meta AI" w/10 EPS) OR ("Meta AI" w/10 "earnings per share")	22	35
((Matthew w/3 Klam) OR (Junot w/3 Diaz) OR (Junot w/3 Díaz) OR (Lysa w/3 Terkeurst)) w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	22	24
(Llama w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	20	51
("Meta AI" w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	13	20

(llama w/10 license w/10 pric*) OR (llama w/10 license w/10 fee) OR (llama w/10 license w/10 charge) OR (llama w/10 license w/10 royalty)	12	25
((metaAI w/5 Facebook) OR (metaAI w/5 Instagram) OR (metaAI w/5 Reels) OR (metaAI w/5 whatsapp) OR (metaAI w/5 Threads) OR (metaAI w/5 oculus) OR (metaAI w/5 metaverse) OR (metaAI w/5 orion) OR (metaAI w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization)	12	12
(eleuther w/20 books3) OR (eleutherAI w/20 books3)	11	13
(book* w/10 long-range) OR (book* w/10 "context model*")	5	57
eye w/20 books3	5	13
(MetaAI w/10 budget) OR (MetaAI w/10 "operational expenses") OR (MetaAI w/10 "gross margin") OR (MetaAI w/10 "operating margin") OR (MetaAI w/10 profit) OR ("Met AI" w/10 EPS) OR (MetaAI w/10 "earnings per share")	3	25
("Sandman Slim" OR Bedwetter OR Ararat OR "Aloha from Hell" OR "Everything Box" OR "Kill the Dead" OR "Perdition Score" OR "Wrong Dead Guy" OR "Dead Ringers") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	1	3
("The Best Yes" OR (Unglued w/10 TerKeurst) OR "Unglued: Making Wise Choices") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	1	2
("After I'm Gone" OR "No Visible Bruises" OR "Brown Girl Dreaming" OR "Water Dancer" OR "We Were Eight Years in Power" OR "Brief Wondrous Life of Oscar Wao") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0

("How it was for me" OR (Less w/10 Greer) OR "Path of Minor Planets" OR "Golden Child" OR "Trying to find Chinatown and Bondage" OR "Trying to find Chinatown & Bondage") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0
("Pandora Room" OR Snowblind OR "Beautiful Struggle" OR (Drown w/10 Diaz) OR (Drown w/10 Díaz) OR "Confessions of Max Tivoli" OR "M. Butterfly" OR "Who is Rich?") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0
("Sam, the cat and other stories" OR "Sam, the cat & other stories" OR "In a Strange City" OR "Lady in the Lake") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0
((Embraced w/20 TerKeurst) OR (Embraced w/5 Devotions) OR "It's Not Supposed to Be this way Study Guide" OR "Uninvited – Study Guide" OR "Made to Crave Devotional") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0
((Feathers w/10 Woodson) OR "Harbor Me" OR "If You Come Softly" OR "Miracle's Boys" OR "Red at the Bone" OR (Uninvited w/20 TerKeurst) OR "Capture His Heart") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0

((Sunburn w/10 Lippman) OR "What the dead know" OR "Wilde Lake" OR "After Tupac" OR "Another Brooklyn" OR (Behind w/10 Woodson) OR "Beneath a Meth Moon") w/15 (llama OR llm OR llms OR "large language model" OR "large language models" OR pre-train OR pre-trains OR pretrained OR pre-training OR dataset OR datasets OR "data set" OR "training data" OR "books 3" OR books3 OR bg3 OR "the Pile" OR reproduc* OR memoriz* OR copy OR copies)	0	0
(MetaAI w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	0	0

Search Terms for Non-email and Workplace Chat

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
Books3 OR b3g OR Gutenberg OR pre-train OR pre-training OR pre-trains OR pre-trained OR pretrain OR pretrains OR pretraining OR pretrained OR dataset OR datasets OR "data set" OR "data sets" OR "training data" OR annotation OR annotations OR Llama OR Genesis OR LLM OR LLMs OR "large language model" OR "large language models" OR Eleuther OR EleutherAI	41,495	43,332
(Copyright* w/15 compensat*) OR (Copyright* w/15 consent*) OR (Copyright* w/15 licens*) OR (copyright* w/15 credit*) OR (copyright* w/15 attribut*) OR (author* w/15 compensat*) OR (author* w/15 consent*) OR (author* w/15 licens*) OR (author* w/15 credit*) OR (author* w/15 attribut*)	5,172	5,630
██ ██ ██ ██ ██ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	1,836	2,042
(mitigat* w/20 memoriz*) OR (mitigat* w/20 copyright) OR (mitigat* w/20 IP) OR (mitigat* w/20 "intellectual property")	1,560	1,615
██ ██ ██ ██ ██ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	1,291	1,418
libgen OR "library genesis" OR "anna's archive" OR "annas archive" OR bibliotik OR zlibrary OR z-library	810	836
██ AND (licens* OR contract* OR agreement* OR negotiat* OR partner*)	747	770
((("meta AI" w/5 Facebook) OR ("meta AI" w/5 Instagram) OR ("meta AI" w/5 Reels) OR ("meta AI" w/5 whatsapp) OR ("meta AI" w/5 Threads) OR ("meta AI" w/5 oculus) OR ("meta AI" w/5 metaverse) OR ("meta AI" w/5 orion) OR ("meta AI" w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization))	527	950
((book* w/20 storytelling) OR (book* w/20 narrative) OR (book* w/20 value) OR (book* w/20 benefit)) AND (dataset OR "training data")	442	446
"theorem proving" AND (dataset OR "training data")	421	425

(Llama w/10 budget) OR (Llama w/10 "operational expenses") OR (Llama w/10 "gross margin") OR (Llama w/10 "operating margin") OR (Llama w/10 profit) OR (Llama w/10 EPS) OR (Llama w/10 "earnings per share")	413	413
torrent AND (dataset OR "training data")	209	217
((Llama w/5 Facebook) OR (Llama w/5 Instagram) OR (Llama w/5 Reels) OR (Llama w/5 whatsapp) OR (Llama w/5 Threads) OR (Llama w/5 oculus) OR (Llama w/5 metaverse) OR (Llama w/5 orion) OR (Llama w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization)	208	731
(licens* w/3 books) OR (licens* w/3 textbooks) OR (licens* w/3 novels) OR (licens* w/3 literature)	194	240
"publicly available datasets" AND llama	189	707
(litigation OR lawsuit OR "fair use") w/20 (Kadrey OR Silverman OR Golden OR Woodson OR Greer OR Snyder OR Hwang OR Coates OR Lippman OR Klam OR Díaz OR Díaz OR Terkeurst OR Farnsworth)	99	252
("Meta AI" w/10 budget) OR ("Meta AI" w/10 "operational expenses") OR ("Meta AI" w/10 "gross margin") OR ("Meta AI" w/10 "operating margin") OR ("Meta AI" w/10 profit) OR ("Meta AI" w/10 EPS) OR ("Meta AI" w/10 "earnings per share")	96	96
((metaAI w/5 Facebook) OR (metaAI w/5 Instagram) OR (metaAI w/5 Reels) OR (metaAI w/5 whatsapp) OR (metaAI w/5 Threads) OR (metaAI w/5 oculus) OR (metaAI w/5 metaverse) OR (metaAI w/5 orion) OR (metaAI w/5 rayban)) AND (integrate OR incorporate OR augment OR monetize OR monetization)	88	90
(Llama w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	43	46
(christopher w/3 farnsworth) OR "blood oath" OR "president's vampire" OR "red, white, and blood" OR "red, white, & blood" OR "a nathaniel cade story" OR "the burning men" OR "eternal world" OR killfile OR "Deep state: a nathaniel cade story" OR ("deep state" AND farnsworth) OR (flashmob AND farnsworth)	35	35
(MetaAI w/10 budget) OR (MetaAI w/10 "operational expenses") OR (MetaAI w/10 "gross margin") OR (MetaAI w/10 "operating margin") OR (MetaAI w/10 profit) OR ("Met AI" w/10 EPS) OR (MetaAI w/10 "earnings per share")	30	40
(book* w/10 long-range) OR (book* w/10 "context model*")	22	22
("Meta AI" w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	16	16
(eleuther w/20 books3) OR (eleutherAI w/20 books3)	12	12

(llama w/10 license w/10 pric*) OR (llama w/10 license w/10 fee) OR (llama w/10 license w/10 charge) OR (llama w/10 license w/10 royalty)	11	142
(MetaAI w/15 stock) AND (price OR valuation OR increase OR impact OR decrease)	8	8
eye w/20 books3	0	0

Search Terms for Mark Zuckerberg (WhatsApp)

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
Llama	599	599
Books	112	112
Textbooks	3	3
Books3	0	0
B3g	0	0
Libgen	0	0
"library genesis"	0	0
Publisher*	29	29
Licens*	195	195
Copyright	13	13
IP	43	43

Search Terms for Mark Zuckerberg (SMS and MMS)

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
Llama	0	0
Books	3	3
Textbooks	0	0
Books3	0	0
B3g	0	0
Libgen	0	0
"library genesis"	0	0
Publisher*	0	0
Licens*	0	0
Copyright	0	0
IP	4	4

Search Terms for Joelle Pineau (WhatsApp)

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
Llama	62	83
Books	1	2
Textbooks	0	0
Books3	0	0
B3g	0	0
Libgen	0	0
"library genesis"	0	0
Publisher*	0	0
Licens*	15	21
Copyright	0	0
IP	2	4

Search Terms for Ahmad Al-Dahle (WhatsApp)

Search Term	Total Documents Hit By Term	Total Documents Hit by Term Including Families
Llama	227	227
Books	15	15
Textbooks	0	0
Books3	0	0
B3g	0	0
Libgen	0	0
"library genesis"	0	0
Publisher*	1	1
Licens*	54	54
Copyright	3	3
IP	12	12